

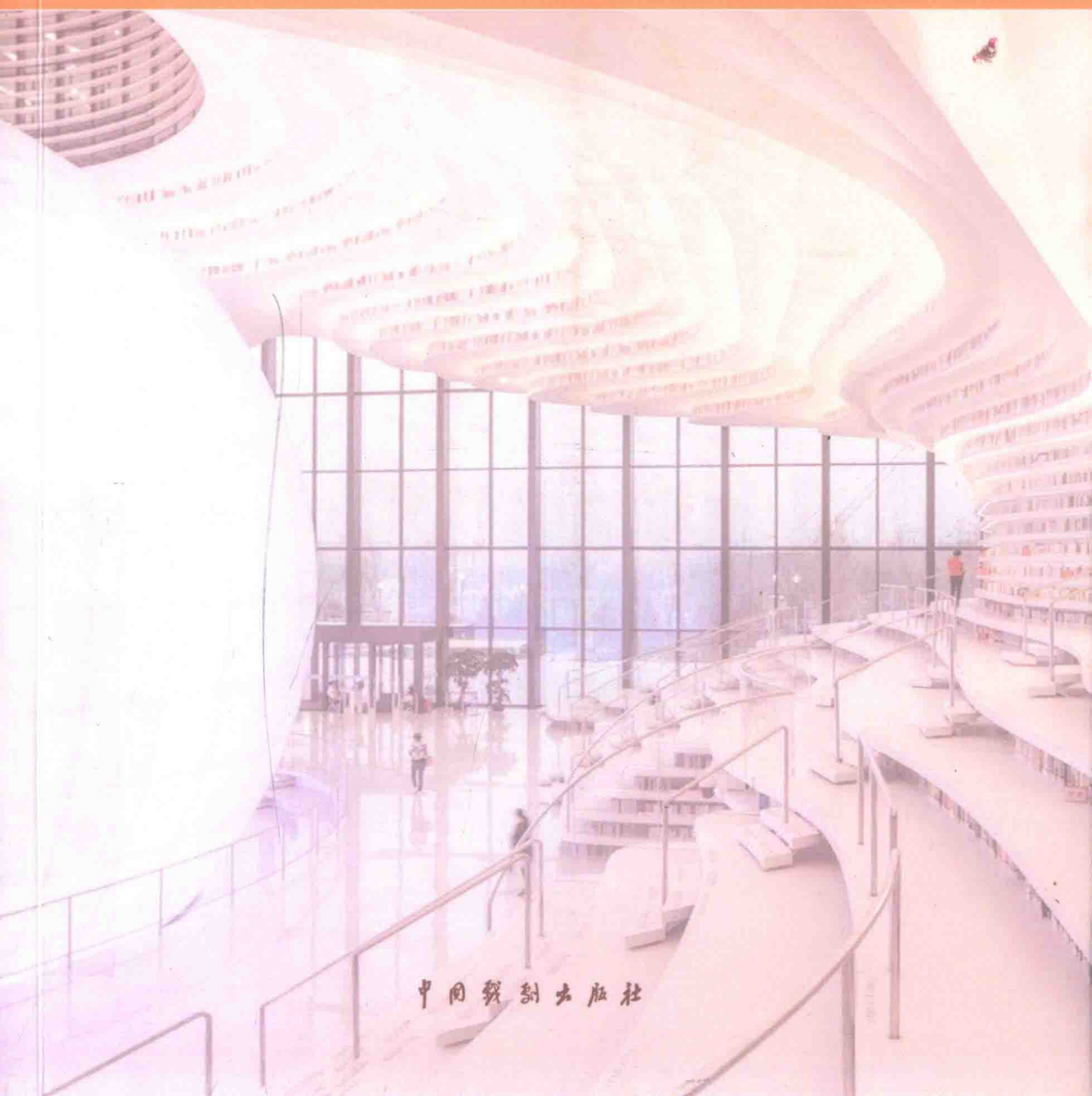
现代图书馆

XIANDAI TUSHUGUAN

图像数据资源建设概论

TUXIANG SHUJU ZIYUAN JIANSHE GAILUN

刘晓辉 著



中国轻工业出版社

XIANDAI TUSHUGUAN

现代图书馆

图像数据资源建设概论

TUXIANG SHUJU ZIYUAN JIANSHE GAILUN

刘晓辉 著

中国文联出版社

图书在版编目 (CIP) 数据

现代图书馆图像数据资源建设概论 / 刘晓辉著.

—北京: 中国戏剧出版社, 2018. 6

ISBN 978-7-104-04672-1

I. ①现… II. ①刘… III. ①现代图书馆—图像数据库—研究 IV. ①G250.74

中国版本图书馆CIP数据核字 (2018) 第104195号

现代图书馆图像数据资源建设概论

责任编辑: 黄艳华

项目统筹: 张 霞

责任印制: 冯志强

出版发行: 中国戏剧出版社

出 版 人: 樊国宾

社 址: 北京市西城区天宁寺前街 2 号国家音乐产业基地 L 座

网 址: www.theatrebook.cn

电 话: 010-63381560 (发行部) 010-63385980 (总编室)

传 真: 010-63383910 (发行部)

读者服务: 010-63387810

邮购地址: 北京市西城区天宁寺前街 2 号国家音乐产业基地 L 座 (100055)

印 刷: 北京盛通印刷股份有限公司

开 本: 710mm×1000mm 1/16

印 张: 18

字 数: 223 千字

版 次: 2018 年 6 月 北京第 1 版第 1 次印刷

书 号: ISBN 978-7-104-04672-1

定 价: 98.00 元



版权所有, 违者必究; 如有质量问题, 请与出版社联系。

前 言

随着社会的进步和现代数字网络信息技术的发展,图书馆以图书、期刊文献纸质资料为主要形式的馆藏模式正在悄然发生转变,大量的书刊文献、图像、录音、录像以及实物资料正在以数字的形式被广泛收藏,保存于数据库,并通过数据库发布系统以文档、图像和音视频等方式展现给读者。其中图像在现实生活中已经成为一种被普遍接受的信息载体,成为语言和文字之外的重要信息传递工具,备受重视。此外,近年来,我国图书馆的图像数据库建设发展十分迅速,但是由于主观和客观等诸多方面因素的影响,导致图像数据库的建设也存在一些问题,高品质的管理系统为数不多。因此,为了解决图书馆图像数据资源建设中的相关问题,实现图像数据库高品质的发展之路,有必要对图书馆图像数据资源建设及相关问题进行深入的研究。

本书作者供职于中国艺术研究院图书馆,作为该图书馆图像工作室负责人,在适应信息时代图书馆图像信息现代化管理的需要方面,进行了多年研究和实践。本书基于图像信息的特点,结合图书馆科学的管理体系,从图像信息的采集、整理、保存、用户服务以及图像部门的业务和人力资源管理等诸多方面进行了深入的探讨,并对高校、

专业等图书馆特色图像数据库建设以及图书馆和图像数据资源建设的未来发展问题做了研究。同时，作者结合已有的科研成果，提出了对图书馆图像数据资源建设的理念，以供图书馆界同仁参考借鉴，并希望得到反馈和建议，以使图书馆图像数据资源能够更好地满足广大用户深层次多样化的需求。

作者

2018 年 3 月

目 录

前 言 / 1

第一章 现代图书馆数据库建设概述 / 001

第一节 数据库概述 / 001

第二节 我国图书馆数据库建设的现状 / 005

第三节 图书馆数据库建设的重要性和可行性 / 007

第四节 图书馆数据库建设的主要方向 / 009

第五节 图书馆数据库建设的实施 / 012

第二章 我国图书馆图像数据库概况 / 017

第一节 图书馆建设图像数据库的意义 / 017

第二节 图书馆图像数据库建设现状 / 020

第三章 图书馆图像数据库建设 / 024

第一节 图书馆图像数据库建设的基本原则 / 024

第二节 古籍图像数据库建设 / 027

第三节 地方图像数据库建设 / 036

第四节 高校图书馆特色图像数据库建设 / 041

第五节 专业图书馆图像数据库建设 / 052

第四章 图书馆图像数据资源的采集 / 063

第一节 图书馆图像数据资源采集机制分析 / 064

第二节 图书馆图像数据资源采集策略和途径 / 066

第三节 图书馆图像信息的拍摄 / 070

第五章 图书馆图像资源的数字加工和存储 / 097

第一节 图像资源数字加工标准的原则 / 097

第二节 图像资源数字加工标准的应用 / 101

第三节 图书馆图像数字化加工 / 104

第四节 图书馆图像数字资源的存储 / 115

第六章 图书馆图像数据资源的检索 / 123

第一节 数字图书馆中基于内容的图像检索技术 / 123

第二节 数字图书馆中基于元数据的图像检索技术 / 126

第三节 数字图书馆中数字图像处理技术 / 130

第四节	数字图书馆图像资源检索	/ 136
-----	-------------	-------

第七章 图书馆图像信息服务 / 139

第一节	图书馆智慧信息服务模式的构建	/ 139
第二节	图书馆个性化信息服务	/ 144
第三节	图书馆特色化信息服务	/ 164
第四节	图书馆学科馆员信息服务	/ 168
第五节	图书馆交互式信息服务的用户激励机制	/ 174
第六节	图书馆图像信息知识服务	/ 177
第七节	加强图像研究促进图像信息增值服务	/ 184

第八章 图书馆图像信息生态管理 / 192

第一节	生态环境下现代图书馆管理概述	/ 192
第二节	生态式管理理念下图书馆管理制度的革新	/ 196
第三节	生态式图书馆的建设	/ 198
第四节	图书馆图像信息资源的可持续性发展	/ 201
第五节	图书馆图像信息的生态管理	/ 206

第九章 图书馆图像馆员人力资源管理 / 214

第一节	概述	/ 214
第二节	图书馆图像馆员人力资源管理的原则	/ 217
第三节	图书馆图像馆员人力资源管理的内容和措施	/ 218
第四节	未来图书馆图像馆员人力资源管理趋势	/ 224

第十章 图书馆图像部门业务管理 / 226

第一节 图像部门的职能属性 / 226

第二节 图像部门业务管理的内容 / 228

第三节 图像部门的岗位设置及职责 / 230

第四节 图像部门的规章制度 / 233

第十一章 图书馆及图像数据资源建设发展思考 / 244

第一节 图书馆建设发展趋势分析 / 244

第二节 图书馆数据资源建设的发展方向 / 247

第三节 图书馆图像数据资源建设的发展趋势 / 249

附录一 中国图书馆员职业道德准则（试行） / 253

附录二 图书馆服务宣言 / 254

附录三 中华人民共和国著作权法 / 255

参考文献 / 272

后 记 / 277

第一章 现代图书馆数据库建设概述

第一节 数据库概述

数据库 (Database) 是按照数据结构来组织、存储和管理数据的仓库, 它产生于 20 世纪 50 年代, 随着信息技术和市场的发展, 20 世纪 90 年代以后, 转变成用户所需要的各种数据管理的方式。数据库有很多种类, 从最简单的存储有各种数据的表格到能够进行海量数据存储的大型数据库系统都在各个方面得到了广泛的应用。

数据管理技术的发展与计算机硬件 (主要是外部存储器) 系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了以下四个阶段: 人工管理阶段、文件系统阶段、数据库阶段和高级数据库技术阶段。

1950 年雷明顿·兰德公司 (Remington Rand Inc.) 生产出第一台商用计算机 Univac I, 推出了一种一秒钟可以输入数百条记录的磁带驱动器, 从而引发了数据管理的革命。1956 年国际商业机器公司 (International Business Machines Corp., 简称 IBM) 生产出第一个磁盘驱

驱动器——the Model 305 RAMAC (Random Access Method of Accounting and Control, 可利用硬盘随机存储和管理数据)。此驱动器有 50 个盘片, 每个盘片直径是 2 英尺, 可以储存 5 兆 (MB) 的数据。使用磁盘最大的好处是可以随机存取数据, 而穿孔卡片和磁带只能顺序存取数据。

最早出现的数据库管理系统 DBMS (Databasae Management System), 是美国通用电气公司 (General Electric Co) 查理斯·巴克曼 (Charles Bachman) 等人在 1961 年开发的 IDS (Integrated Data Store)。1964 年他又成功地开发出世界上第一个网状 DBMS, 也即第一个数据库管理系统——集成数据存储 IDS (Integrated Data Store), 奠定了网状数据库的基础, 并在当时得到了广泛的发行和应用。之后, 通用电气公司一个客户——百路驰化学公司 (BF Goodrich Chemical) 重写了整个系统, 并将重写后的系统命名为集成数据管理系统 (IDMS)。

层次型数据库管理系统 (DBMS) 是紧随网络型数据库而出现的, 最著名、最典型的层次数据库系统是 IBM 公司在 1968 年开发的 IMS (Information Management System), 一种适合其主机的层次数据库。这是 IBM 公司研制的最早的大型数据库系统程序产品。从 20 世纪 60 年代末其产生算起, 如今已经发展到 IMSV6, 提供群集、N 路数据共享、消息队列共享等先进性能的支持。

1969 年埃德加·弗兰克·科德 (Edgar Frank Codd), 通常被称为“泰德” (Ted), 发明了关系数据库 (SQL)。1970 年, 时为 IBM 的研究员他在《美国计算机协会通讯》 (Communication of the ACM) 上发表了一篇名为《大型共享数据库数据的关系模型》 (A Relational Model of Data for Large Shared Data Banks) 的论文, 提出了关系模型的概念, 奠定了关系模型的理论基础。这篇论文被普遍认为是数据库系统历史上具有划时代意义的里程碑。1973 年加州大学伯克利分校的迈克尔·斯通布雷克

(Michael Stonebraker) 和王佑曾 (Eugene Wong) 利用 R 系统 (System R) 已发布的信息开始研发自己的交互式图形和检索数据库系统 (Interactive Graphic and retrieval system, 简称 Ingres)。他们开发的 Ingres 项目最后由甲骨文公司 (Oracle)、英格来 (译音) 公司 (Ingres) 以及硅谷的其他厂商所商品化。后来, System R 和 Ingres 系统双双获得美国计算机协会 (ACM) 1988 年颁发的“软件系统奖”。

1976 年霍尼韦尔公司 (Honeywell) 开发了第一个商用关系数据库系统——Multics Relational Data Store。关系型数据库系统以关系代数为坚实的理论基础, 经过几十年的发展和实际应用, 技术越来越成熟和完善。其代表产品有甲骨文公司的 MySQL、IBM 公司的 DB2、微软公司的 MS SQL Server 以及 Informix、ADABAS D, 等等。

MySQL 是最流行的关系型数据库管理系统, 在网页 (web) 应用方面 MySQL 是最好的关系型数据库管理系统 (Relational Database Management System) 应用软件之一。MySQL 是一种关系型数据库管理系统, 关系型数据库将数据保存在不同的表中, 而不是将所有数据放在一个大仓库内, 这样就增加了速度并提高了灵活性。MySQL 所使用的 SQL 语言是用于访问数据库的最常用标准化语言。MySQL 软件采用了双授权政策, 它分为社区版和商业版, 由于其体积小、速度快、总体拥有成本低, 尤其是开放源码这一特点, 一般中小型图书馆的开发都选择 MySQL 作为资源数据库。

NoSQL, 泛指非关系型的数据库, 是近十年来逐渐兴起的数据库系统。因其基于键值对, 可以想象成表中的主键和值的对应关系, 而且不需要经过 SQL 层的解析, 所以性能非常强; 另外, 因为基于键值对, 数据之间没有耦合性, 所以具有较强的扩展性等优点。随着互联网 web2.0 网站的兴起, 传统的关系数据库在应付 web2.0 网站, 特别是超大规模和

高并发的 SNS 类型的 web2.0 纯动态网站已经显得力不从心，暴露了很多难以克服的问题，而非关系型的数据库则由于其本身的特点得到了非常迅速的发展。目前两种类型的数据库逐渐相互融合，优势互补，为大数据时代的数据库建设充当强有力的基石。

MongoDB 是一个介于关系数据库和非关系数据库之间的产品，是非关系数据库当中功能最丰富、最像关系数据库的，支持的数据结构较为松散，因此可以存储复杂的数据类型。它还是目前在 IT 行业中非常流行的一种分布式文件存储的数据库，旨在为网页（web）应用提供可扩展的高性能数据库存储解决方案，且操作简单、源代码公开免费，可随时下载使用。其优点包括：①支持的查询语言非常强大。②内置的水平扩展机制提供了从百万到十亿级别的数据量处理能力，完全可以满足社交媒体数据存储要求。③面向集合存储，易存储对象类型的数据，包括视频等大型对象。④模式自由，对于存储在其中的数据文件，不需要知道其任何结构意义。

目前，图书馆通常采用 SQL 数据库，并结合各馆资源特点、馆员结构和用户需求等具体情况设计管理系统软件，采取国际流行的标准技术规范进行数据加工，建设自己的馆藏数据资源管理体系。SQL 数据库的优点主要有：可以用 SQL 语句便捷地在一个表以及多个表之间做非常复杂的数据查询；可以实现对于安全性能很高的数据访问要求。但随着大数据时代的海量存储，SQL 数据库开始逐渐融合 NoSQL 数据库中的优势性能，弥补自身存在的缺陷。可以想见，未来图书馆数据库建设将不断融入最新科技成果，沿着健康的方向发展。

第二节 我国图书馆数据库建设的现状

一、起步较迟、发展迅速

我国数据库建设已有近 30 年的历史,其发展过程大约可以分为三个阶段:一为起步阶段(1975~1979),主要是引进、学习、借鉴国外数据库的理论成果;二为发展阶段(1980~1993),主要是研究和自建中文数据库。据统计,20 世纪 90 年代初我国自建的不同规模的数据库就有 806 个。三为成熟、实用及飞速发展阶段(1993 年至今),以 1993 年 2 月我国第一家专业数据库公司——“万方数据公司”的正式成立为标志。到 1995 年,我国自建的数据库已有 1038 个,容量在 10MB 以上的增加至 67%,其中 100MB 以上的占 25%;1999 年 9 月,我国网上信息资源和应用数据库达到 4400 多个,而据 2005 年 2 月中国互联网信息中心的调查,我国在线数据库已达 45598 个,其发展速度之快,令人惊叹。目前,最有代表性的书目数据库有国家图书馆开发的《中国国家书目数据库》和中科院的《西文期刊联合目录数据库》,全文数字期刊有《中国知识资源总库》、清华《学术期刊数据库》、重庆维普《中文科技期刊数据库》、人大《报刊复印资料数据库》、万方公司《万方数据库资源系统》等,全文数字图书有书生公司的《书生之家》、超星公司的《超星数字图书馆》、数图公司的《数字图书》等,都已具备了相当的规模。在学位论文方面有《中国学位论文数据库(CDDB)》及中国高等教育文献保障系统(CALIS)的《高校论文数据库》。在引进数据库方面,购买了一大批优秀的外文数据库,如联机计算机图书馆中心(OCLC)《第一检索(First search)数据库系统》及《英国物理学会期刊数据库》(IOP)、美国《科学引文索引》(SCI)等,极大地充实了我国图书馆数据库建设的规模。

二、条块分离、各自为政

长期以来,我国的文献服务机构由三大服务群体组成,即科技专业系统(主要是中科院图书馆系统和情报所系统)、高校图书馆系统和公共图书馆系统。各服务群体由于隶属关系和管理体制的不同,相互之间缺乏横向联系和整体协调,具体表现在数据库建设上是条块分割、各自为政,本位主义严重,以致于所建成的数据库往往结构单一、规模小、专业面狭窄,且大多只能是自用数据库,根本无法上网使用。而对于那些专门的数据库开发商来说,由于国家对其缺乏宏观调控,既没有一个负责对数据库建设实行整体规划,协调数据库建设多方参与者之间利益的权威性机构,也缺少数据库产业内部行业自律机制的行业协会,这必然造成各数据库开发商各自为政、盲目建库的局面。但随着互联网的普及和网络技术的成熟,三大系统之间实现网络互联和资源互动,已成为今天信息资源社会共享发展的必然趋势,这就暴露了我国图书馆数据库建设“单打独斗”已经成为数据库产业发展的严重桎梏。

三、标准不一,难以共享

数据库建设的标准化、规范化是实现信息资源共建共享和文献检索自动化的重要基础和前提之一。数据库的标准化主要分为两个方面:一是数据库管理系统的标准化;二是数据库资源的标准化。多年来,各图书馆或数据库开发商们由于缺乏统一的管理机构,各自为政,各行其是,自由发展,在数据库建设的标准化和规范化方面处于比较混乱的状态。具体表现在:基于数据库管理系统的检索系统、检索软件和操作系统多种多样;数据库的格式、字段不一;数据的标引、分编、检索点选取没有统一的标准和严格的质量控制。由此而造成的结果是

数据库的兼容性和互用性差,原始数据处理不完备、不准确、不规范、不统一,从而严重影响了数据库的利用和共享,限制了数据库作用的发挥。

第三节 图书馆数据库建设的重要性和可行性

随着现代信息网络技术的飞速发展,信息的内容和数量越来越丰富,人们获取有用信息的途径将不再局限于在图书馆查阅有形的馆藏资料,还可以直接到互联网上寻找自己所需要的信息。因此,传统意义上的图书馆已不能完全适应当前网络环境下广大用户的需要,图书馆只有逐步向数字图书馆的方向发展,才能更好地满足用户对信息的需求。现代图书馆将是集实体资源、数字资源、自动化和网络化于一体的信息集散机构。数据库建设既是实现数字图书馆的基础,也是馆藏资源建设的重要组成部分。图书馆建设高质量的数据库,让用户能够更加便捷、有效和准确的检索到所需信息资源,是大势所趋。

一、适应现代信息社会发展的需要

进入 20 世纪 50 年代以后,计算机技术逐步应用于图书馆信息管理系统之中,促进了图书馆馆藏信息的收集、加工、保存、传播和利用;大容量磁盘存储技术和光盘存储技术的问世,促进了数据库建设在图书馆的广泛开展;通信技术的发展,图书馆可以借助通信网络进行信息传递,网上信息检索、多媒体信息的传递等,使得数据库的广泛利用成为可能,促进了数据库的远程访问和高效利用;多媒体技术的发展,促进

了文字、图形、图像、音频、视频和动画等的交互信息处理，丰富了数据库内涵，也在客观上推动了数据库的快速发展。

二、体现图书馆信息资源建设的发展方向

长期以来，作为重要的信息资源中心，图书馆一直都以纸本资源为主。随着信息技术的发展，图书馆未来的信息资源建设将主要集中于数字化资源的方向。数字化资源与传统信息资源相比，有存贮容量大、保存时间长、检索速度快、使用方便等优势，而且伴随着数字图书馆的发展，数字化资源不受传统图书馆客观条件的限制。图书馆通过对各种信息资源进行综合分析处理，提取主要的信息要素，按一定的规范模式存储到计算机中，形成内容完备、检索便利、面向社会服务的文献数据集合，进而可以形成集咨询、检索、服务等功能于一体的文献信息服务系统。

三、图书馆建设数据库的优势

目前我国大部分图书馆均已接入了网络，图书馆数据库建设的硬件条件已经基本具备，因此首先具备了硬件设备上的优势。其次，许多图书馆经过长达几十年的积累，拥有大量的馆藏图书，在拥有丰富的文献资源的同时还拥有专业的特藏，从而为数据库的建设打下良好的基础，这就具有了信息资源的优势。再次，现代网络技术的飞速发展促使图书馆的服务逐步走向自动化，使其在建设数据库时具备了管理和服务上的优势。最后，也可以说是最重要的一点，那就是人才优势，因为部分图书馆依托高校、科研院所，拥有非常多的专家队伍，具有相当可观、优质可靠的人才优势。下文便以高校、专业科研院所图书馆为例，对图书馆数据库建设的主要方向加以简单介绍和分析。